

DATA ANALYSIS PROJECT

9/22/2021

```
library("ggplot2")
library("forcats")
library("RColorBrewer")
library("dplyr")
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

```
library("knitr")
```

```
Depression <- read.table("/Users/luciabravo/Desktop/math_130/data/depress_081217.txt", header=TRUE, sep=",")
```

DATA:

My project will be based on the Depression data set. This data set contains 294 participants from the Los Angeles County. The data obtained from the participants are results from interviews that were done on each participant. The questions asked in the interviews regarded the person's lifestyle such as their marital status, education, income, physical health, sex, employee status, age, religion, cases, drinking habits, and beddays. After considering the answers of every participant, they are then given a score based on depression scale that is between 1-60, 60 being the highest.

I am specifically interested in looking at the correlation between people's 'drinking habits' and their level in the depression scale, 'cesd.' My anticipation is that there will be a higher level of depression for those who consume alcohol on a regular bases. I am also interested in seeing the correlation between a person's 'chronic illness' and their level in the depression scale, 'cesd.' I anticipate that those who have recently had a chronic illness will have a higher level of depression.

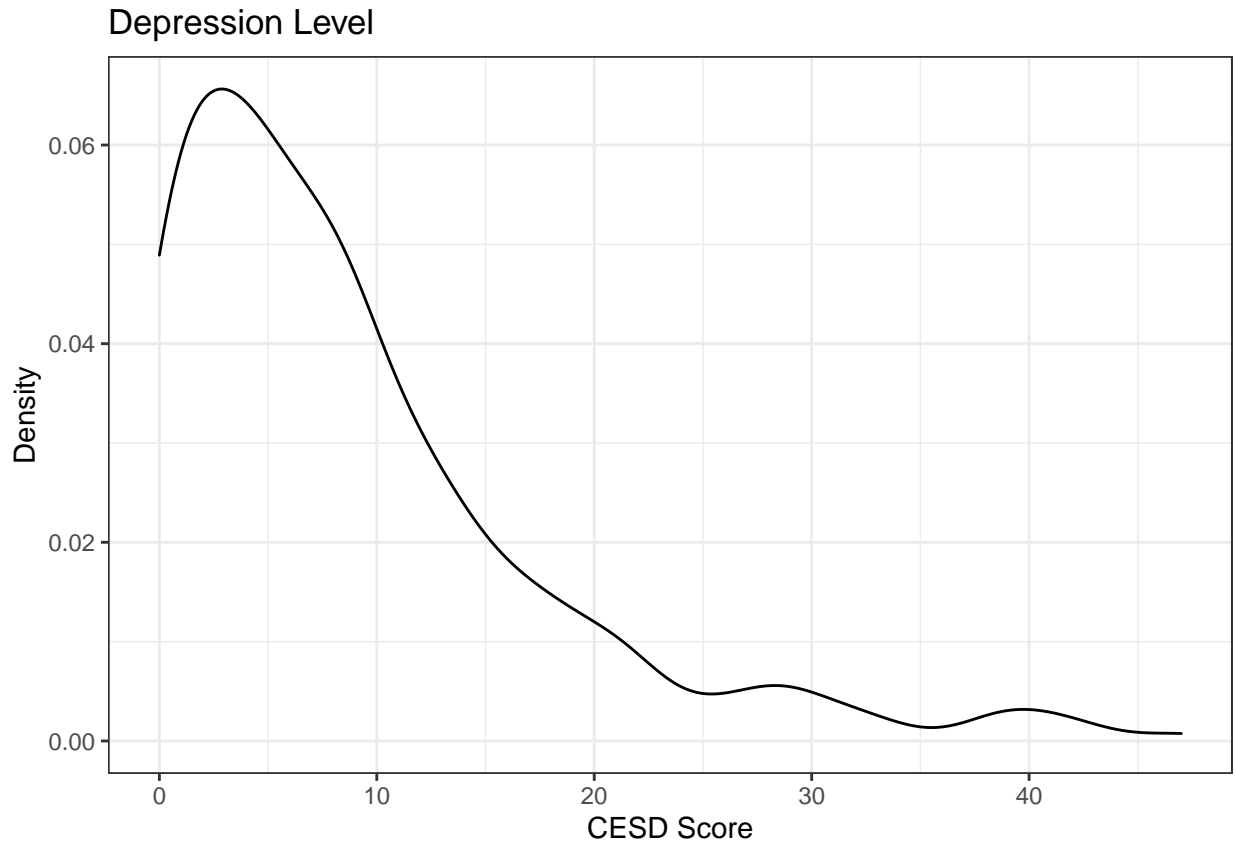
UNIVARIATE DESCRIPTION: 1. CESD VARIABLE: The data shown below is the summary statistics regarding the 'cesd' variable. This variable describes the depression level that each participant has been ranked due to their depression related answers. Upon the 294 participants, the minimum score for this data set was 0, the median was 8.884, and the maximum was 47.

```
summary(Depression$cesd)
```

```
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.000   3.000   7.000   8.884  12.000  47.000
```

The plot below is a density plot of the 'cesd' variable mentioned above. On the y-axis is the density and on the x-axis is the 'cesd' score from each participant. Considering that the graph is skewed to the right, that indicates that the great majority of the participants scored a relatively low score in the depression scale.

```
ggplot(Depression, aes(x=cesd)) + geom_density() + ggtitle("Depression Level") + xlab("CESD Score") + ylab("Density")
```



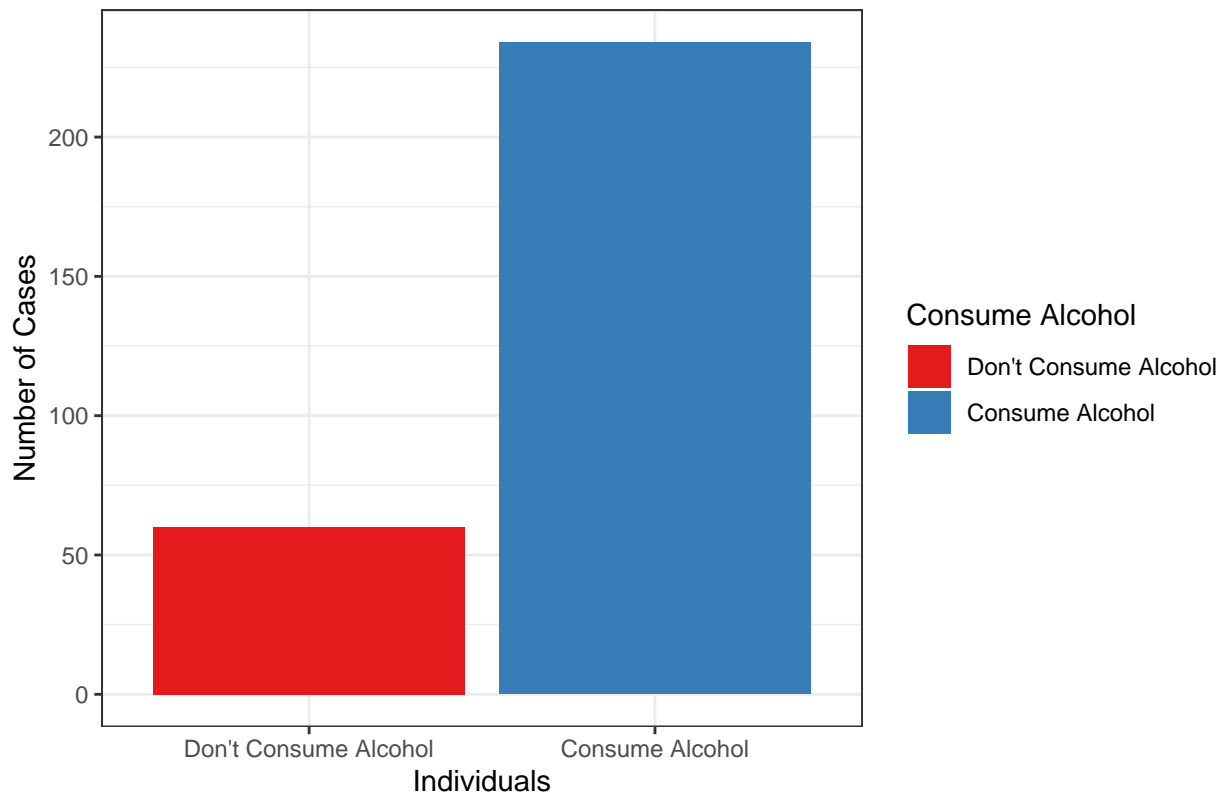
DRINKING HABITS VARIABLE The table and the bar graph below are regarding the 'drinking' variable in this data. The table shows the amount of participants that consume alcohol and those who do not. Out of the 294 participants, 243 individuals consume alcohol and 60 individuals do not. These results are also represented in the bar graph below. On the y-axis is the number of cases and on the x-axis are the individuals in the data. The red bar represent those who do not consume alcohol and the blue bar represent those who do consume alcohol.

```
Depression$drink <- factor(Depression$drink, labels=c("Don't Consume Alcohol", "Consume Alcohol"))
table(Depression$drink, useNA="always")
```

```
##
## Don't Consume Alcohol      Consume Alcohol      <NA>
##                60                234                0
```

```
ggplot(Depression, aes(x=drink, fill=drink)) + theme_bw() + geom_bar() + scale_fill_brewer(palette="Set1")
```

Participants Alcohol Consumption



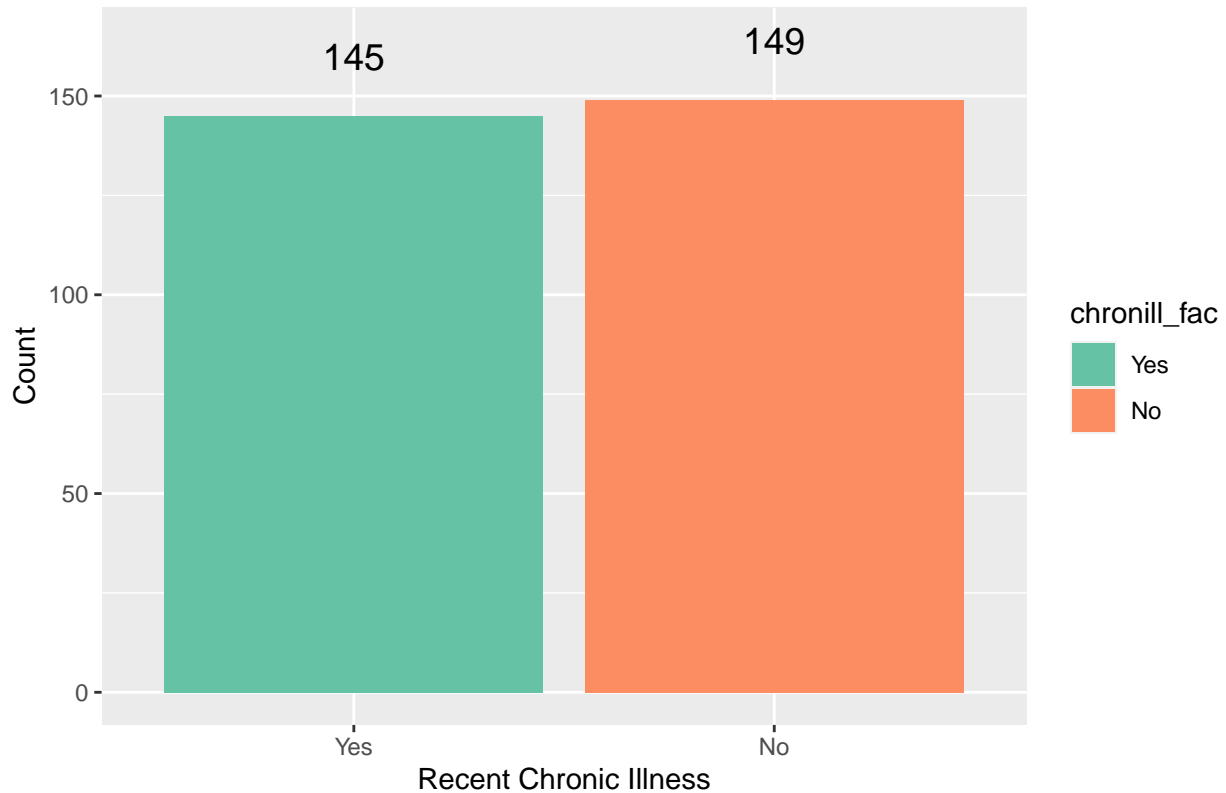
2. CHRONIC ILLNESS VARIABLE: The table and the bar graph below are regarding the Chronic Illness Variable in this data. The table shows the amount of participants that have experienced a chronic illness recently and those who did not. Out of the 294 participants, 145 individuals have recently experienced having a chronic illness and 149 individuals have not. These results are also represented in the bar graph below. On the y-axis is the count of of the amount of people who have or haven't had a chronic illness recently and on the x-axis are the two possibilities of having a chronic illness or not having a chronic illness. The teal bar represents those who have had a chronic illness recently and the peach bar represents those who have not had a chronic illness.

```
Depression$chronill_fac <- factor(Depression$chronill, labels=c("Yes","No"))  
table(Depression$chronill_fac)
```

```
##  
## Yes No  
## 145 149
```

```
ggplot(Depression, aes(x=chronill_fac, fill=chronill_fac)) + geom_bar(aes(y=..count..)) + ggtitle("Chr
```

Chronic Illness Within Participants



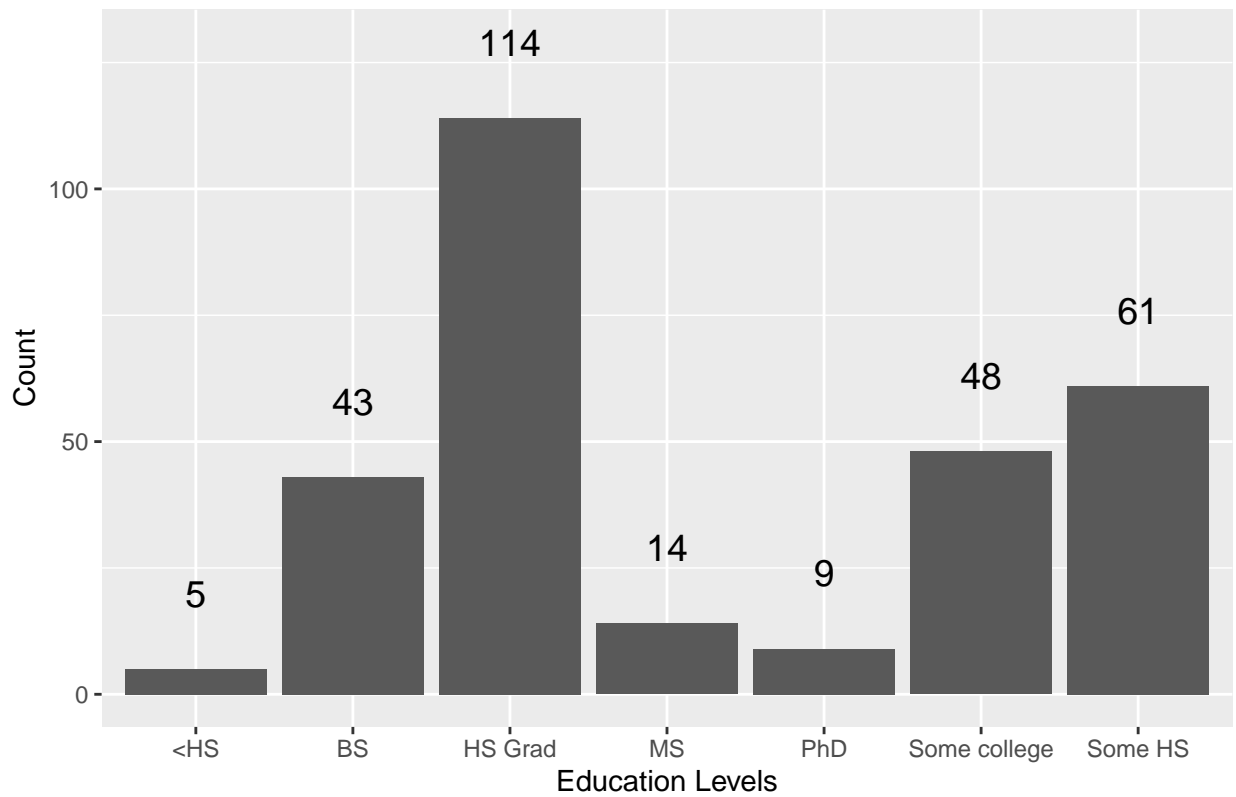
3. EDUCATION VARIABLE: The table and the bar graph below are regarding the Education Variable in this data. The table shows the education level of every participant in the data. Out of the 294 participants, 5 individuals have less than a high school education, 61 had some high school education, 114 graduated high school, 48 has some college education, 43 have graduated college with a BS, 14 have a MS, and 9 individuals have their PhD. These results are also represented in the bar graph below. On the y-axis is the count of the participant’s education level and on the x-axis are the different types of education levels.

```
table(Depression$educat)
```

```
##
##      <HS      BS      HS Grad      MS      PhD Some college
##      5      43      114      14      9      48
##      Some HS
##      61
```

```
ggplot(data=Depression,mapping=aes(x=educat)) + geom_bar(aes(y=..count..)) + ggtitle("Participants Educac")
```

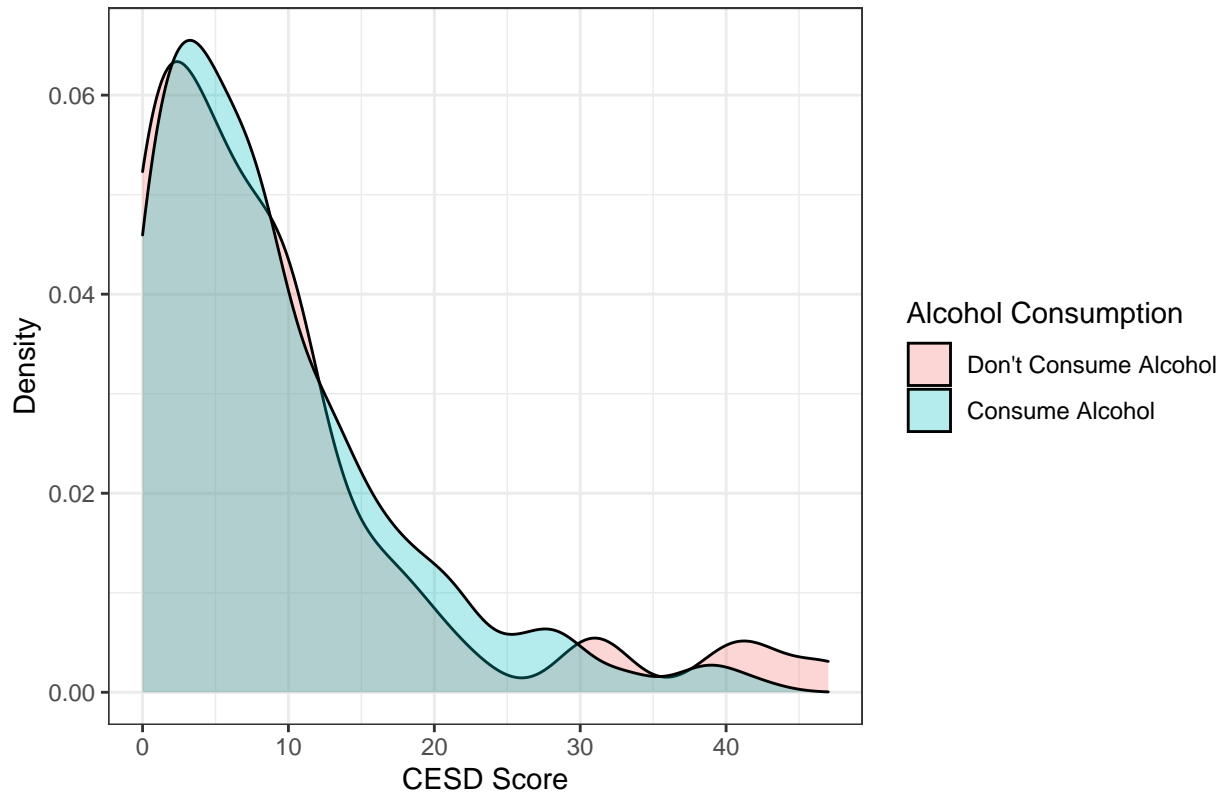
Participants Education



BIVARIATE COMPARISON 1. Comparison between the variables, “Drink” and “cesd” In the graph below, I compared the two variables, ‘cesd’ and ‘drink.’ I wanted to see if there was any sort of correlation between the two variables, considering that alcohol could influence an individuals mental health. However, by the results of the graph, you could see that there isn’t a significant correlation between the two variables as I anticipated.

```
ggplot(Depression, aes(x=cesd, fill=drink)) + geom_density(alpha=.3) + scale_fill_discrete(name="Alcohol")
```

Drinking and CESD



2. Comparison between the variables, "cesd" and "Chronic Illness" In the graph below, I compared the two variables, 'Cesd' and 'Chronic Illness.' I wanted to see if there was any correlation between the two variables, considering that a person's physical health could affect a person's day to day mood. However, by the results of the graph, you could see that there is not any significant correlation of the two data variables.

```
ggplot(Depression, aes(x=cesd, fill=chronill_fac)) + geom_density(alpha=.3) + scale_fill_discrete(name=
```

Recent Chronic Illness and CESD

